From:	PETERSON Jenn L
To:	Eric Blischke/R10/USEPA/US@EPA; Jay Field
Cc:	Burt Shephard/R10/USEPA/US@EPA; Joe Goulet/R10/USEPA/US@EPA; Mary Baker; Robert Neely
Subject:	RE: Bioassay Evaluation
Date:	07/15/2008 04:47 PM

At the risk of putting everyone over the edge on this, I would recommend looking at the March 17, 2006 presentation of the models (esp. the FPM) - *Interpretive Report: Estimating Risks to benthic organisms using predictive models Based on Sediment Toxicity Tests.* Although some analysis is missing, if you look at table 5-3 (reliability parameters) and the maps, I don't think there is justification for dropping lower thresholds for all endpoints (e.g. there isn't support for the statement that they can't distinguish between hit and no hit thresholds). Putting the H. growth argument aside for a moment, it is unclear why predictive problems identified by the LWG for the H. growth endpoint would influence the selection of appropriate thresholds for model development for other endpoints / species.

-Jennifer

-----Original Message-----
From: Blischke.Eric@epamail.epa.gov [mailto:Blischke.Eric@epamail.epa.gov]
Sent: Tuesday, July 15, 2008 4:12 PM
To: Jay Field
Cc: Shephard.Burt@epamail.epa.gov; Goulet.Joe@epamail.epa.gov; Mary Baker; PETERSON Jenn L; Robert Neely
Subject: Re: Bioassay Evaluation

I want to offer up a few points in response to Jay and Jennifer's emails.

First, I do understand that we are using the higher RSET thresholds. However, the LWG has maintained that they cannot build a floating percentile model at these lower thresholds because there is not enough of a difference between the hit and no-hit distributions. The logistic regression model has been developed a little differently and uses a screening concentration to account for the fact that we have a lot of high chemistry concentrations that do not result in a hit. As I mentioned in my email, I believe the empirical data is very good at delineating areas of contamination even at low hit thresholds. However, the government team must accept that fact that it is difficult to distinguish between the hit and no-hit distributions for a specific chemical. If we do not accept this, then it is impossible for us to move forward objectively

Second, Jennifer points out that this analysis has not been presented. I am attaching a summary that was prepared by the LWG last fall. Although I have not developed a detailed, critical review of this information, John indicated that the information presented supports his position.

Third, I disagree with the contention that the empirical and modeled thresholds need to be the same. They are for two different purposes. As I mentioned in number one above, the Hyalella growth endpoint appears to be a sensitive and valuable empirical measure of benthic risk. However, it does not appear to work well in the predictive sense. Because the purpose of the predictive models are to develop (under the latest proposal) probable effect levels, there is no compelling reason for them to be the same.

Fourth, with the new Round 3 data, there are approximately 40 samples that exceed the 40% reduction in growth relative to control. I agree that a larger hit data set is preferable but not if we can not distinguish between the hit and no-hit distribution.

Fifth, as I have stated previously, I have little confidence in the models. If I was developing the

Portland Harbor work plan today, I would not include the models. Rather I would go with the more standardized approach of relying on threshold effect levels to identify areas with low probability of toxic effects, probable effect levels to identify areas with a high probability of toxic effects and bioassays to assess the areas in-between. We have spent way too much time, energy and money discussing these models.

Sixth, in a perfect world, we would know how to optimize model performance and based on this optimization step, we would run the models. However, I am not convinced that anyone knows how to optimize the models. At this point the best we can do is to go forth with something, evaluate model performance in the baseline ecological risk assessment and adjust the models as necessary for the final RI and BRA. This is a key element of the proposal.

Finally, I would like say that I am cognizant of the fact that the LRM and the FPM are different, that what works well for one may not work so well for the other and vise versa, and that as a consequence, we may be talking past each other. As a result, I am willing to entertain different thresholds for the LRM (e.g., 20% difference from control for the Hyalella growth endpoint as we proposed earlier).

Eric

(See attached file: LWG slides for Hyalella growth meeting 100507.pdf)

Jay Field
<Jay.Field@noaa.gov>
07/15/2008 10:23 AM

To: Eric Blischke/R10/USEPA/US@EPA
cc: Burt Shephard/R10/USEPA/US@EPA, Robert Neely <Robert.Neely@noaa.gov>, Jennifer Peterson <PETERSON.Jennifer@deq.state.or.us>, Joe Goulet/R10/USEPA/US@EPA, Mary Baker <Mary.Baker@noaa.gov>
Subject: Re: Bioassay Evaluation

Eric,
I completely agree with Jennifer that using model thresholds that deviate significantly from the empirical thresholds for the purposes of defining risk to the benthic community is difficult to understand or justify.

some additional information, based on the existing models:

using the proposed thresholds for Hyalella and Chironomus, there are 52 (of original 233) samples that

would be classified as hits (32 for Hyalella and 39 for Chironomus). 41 of 52 have an LRM PR_Max>0.6 32 of 52 exceed the FPM based on 80% threshold (Q80>1)

51 samples have a PR_max>0.6 and do not exceed the proposed toxicity thresholds 25 samples have a PR_max>0.7 and do not exceed the proposed toxicity thresholds

43 samples exceed the FPM (80%) and do not exceed the proposed toxicity thresholds

21 samples have a PR_max>0.6 and all survival/growth endpoints have greater than or equal to 80% of control 23 samples have a Q80>1 and all survival/growth endpoints have greater than or equal to 80% of control

I conclude from this that both models identify most of these highly toxic samples. With more reasonable thresholds (80% of control for all endpoints), the number of "false positives" is low for both models. In my view, the best approach is to develop the models using a reasonable conservative thresholds (which has already been done) and adjust probability thresholds for risk management.

Jay Field wrote:
> Eric,
> I assume from item #6 that you are proposing to derive new LRMs based
> on the thresholds you identified. The proposed thresholds, which I
> consider to be extreme, will make the development of reliable logistic

> regression models highly unlikely because of the small number of "hit"

> samples. The basis for the new growth thresholds is not clear, given
> that all of the growth results with less than 80% of control were
> statistically different from the control. If your intent is to
> incorporate risk management decisions into the models, I would
> recommend applying those concepts to model application, not model
> development. Please let me know if I can provide further assistance.
> Jay
>
> [6) The evaluation of the bioassay data for the development of the
> predictive models will be based on the following hit thresholds:
> - Chironomus Growth - 30%
> - Chironomus Mortality - 20%
> - Hyalella Growth - 40%
> - Hyalella Mortality - 20%
>
>
>
> Blischke.Eric@epamail.epa.gov wrote:
>> All,
>>
>> We have been discussion the evaluation of the bioassay at length over
>> the last month or so and off and on for the last four years. In
>> general, I believe that the empirical toxicity test results is our
>> strongest line of evidence for evaluating effects on the benthic
>> community. However, I recognize the need to develop site specific
>> predictive models to assist the evaluations of stations where
sediment
>> toxicity test results are not available. At the heart of this
>> discussion has been the use of the Hyalella growth endpoint in the

>> predictive models.
>>
>> An evaluation of the empirical Hyalella growth data suggests that
>> Hyalella growth at the 10% and 20% difference from control hit
>> thresholds can be used to delineate the extent of contamination at
the
>> Portland Harbor Site. However, LWG representatives have maintained
>> that, on a chemical by chemical basis, it is not possible to see a
>> difference between the hit and no-hit distributions at these levels.
As
>> a result, the Round 2 Report did not consider the Hyalella growth
>> endpoint during the development of the floating point percentile
model
>> (FPM). EPA has maintained that the Hyalella growth endpoint adds
>> valuable information and thus should be included in the model.
>>
>> Regarding the evaluation of empirical data, the LWG has agreed to our
>> recommended approach which was to evaluate all four endpoints
(Hyalella
>> growth and mortality; Chironomus growth and mortality) at the 10, 20
and
>> 30% difference from control level. The LWG sought to address our
>> concern about the use of the Hyalella growth endpoint by proposing
the
>> RSET one-hit/two-hit threshold for use in the predictive models.
This
>> proposal was not endorsed by the project team.
>>
>> On Friday afternoon, Burt and I spoke with John Toll and Helle
Anderson
>> about the evaluation of benthic risk. At the end of the discussion,
we
>> came up with the following approach.
>>
>> 1) Evaluate the empirical toxicity data as we have described - a hit
is
>> a statistically significant difference from control for any of the
four
>> endpoints.
>> 2) Substitute total biomass for the growth endpoint for both the
>> Hyalella and the chironomus tests.
>> 3) Empirical data will be further refined by classifying the
toxicity
>> tests into minor (10%) moderate (20%) and severe effects (30%).
>> 4) For the LRM and FPM, we will pool the growth (biomass) and
mortality
>> endpoints for chironomus and again for Hyalella.
>> 5) Pooling will be based on use of the most sensitive endpoint
(growth
>> or mortality) resulting two LRM and two FPM models.
>> 6) The evaluation of the bioassay data for the development of the

>> predictive models will be based on the following hit thresholds:
>> - Chironomus Growth - 30%
>> - Chironomus Mortality - 20%
>> - Hyalella Growth - 40%
>> - Hyalella Mortality - 20%
>> 7) These thresholds will apply to both the logistic and floating
>> percentile models.
>> 8) The results from these models will be equivalent to site
specific
>> probable effect levels.
>> 9) The draft RI report will present an evaluation of the hit
>> thresholds used in the predictive models. The evaluation will
compare
>> the separation of sediment chemistry distributions at the hit and no
hit
>> stations as a way to assess the utility of using lower hit thresholds
in
>> the predictive models, evaluate the reliability of the predictive
models
>> and make recommendations regarding the optimization of model
>> performance.
>> 10) The model results will be used in the conjunction of other lines
of
>> evidence in the baseline risk assessment and in the development of
PRGs.
>>
>> Although the hit thresholds identified for the predictive models are
>> higher than what we have discussed previously, we will perform
analysis
>> on the back end to assess the utility of using lower thresholds. This
>> analysis will be presented in the draft baseline ecological risk
>> assessment and the hit/no-hit thresholds will be adjusted as
necessary
>> prior to the final BERA. In my view, the hit threshold or thresholds
>> selected for use in the predictive models are for the purpose of
>> optimizing model performance. Due to the large number of sources and
>> source types at the Portland Harbor site, the predictive model
results
>> do not necessarily match up well with the empirical bioassay results.
>> In a perfect world, we would perform the necessary analysis to
determine
>> the optimum hit threshold or thresholds prior to running the model.
>> However, the project schedule does not allow this approach. In any
>> event, the predictive model results are only one line of evidence for
>> evaluating risk to the benthic community and will be weighted
>> accordingly in the baseline ecological risk assessment (BERA). These
>> results will be used along with other lines of evidence (e.g., SQGs,
>> application of benthic tissue TRVs and BSAFs) to identify areas that
>> pose risk to the benthic community and develop sediment cleanup
levels
>> protective of the benthic community.
>>

>> Please let me know if you have any questions. I will cover this at this
>> week's TCT.
>>
>> Thanks, Eric
>>
>>
>>
>

--
Jay Field
Assessment and Restoration Division
Office of Response and Restoration, NOAA
7600 Sand Point Way NE
Seattle, WA 98115-6349
(P) 206-526-6404
(F) 206-526-6865
(E) jay.field@noaa.gov